

# SURF DATA EXCHANGE

Freek Dijkstra, Axel Berg, Mike Kotsur, *et al.*

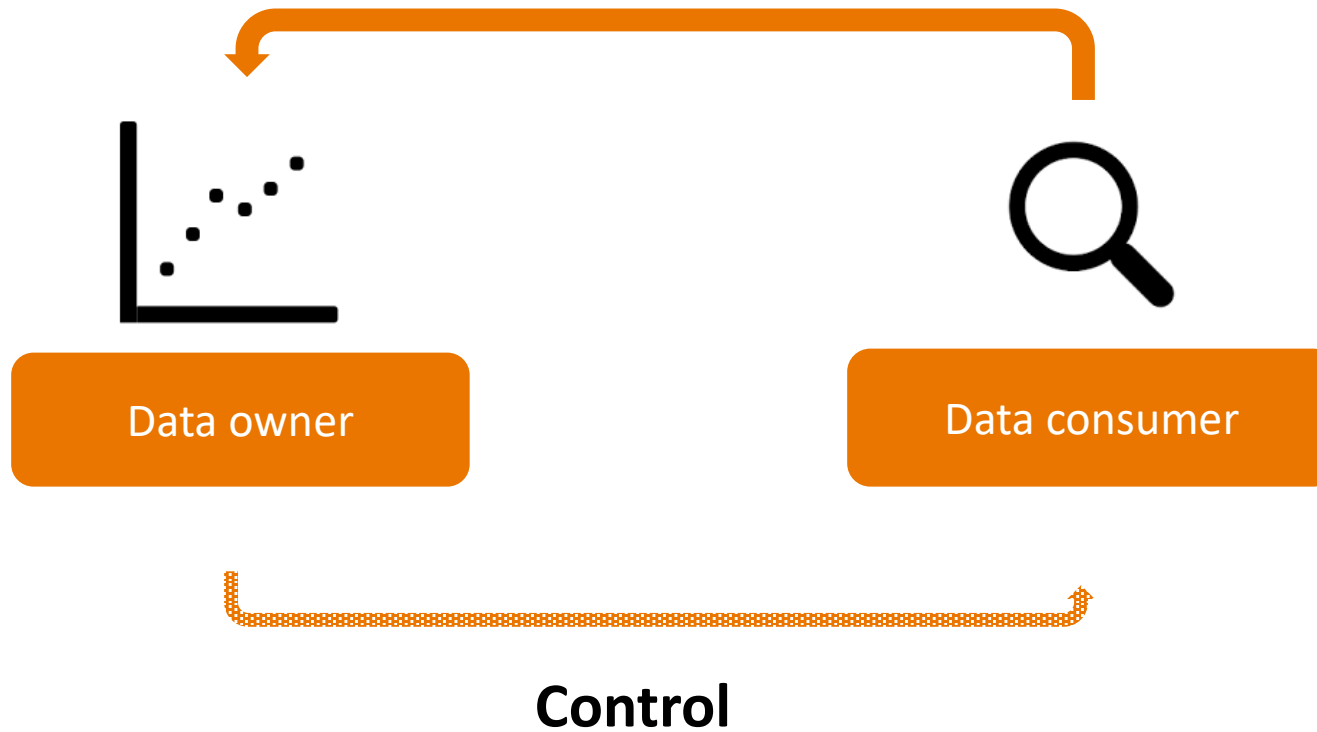
Share data while retaining control and confidentiality of your data

WEBINAR SHARING CONFIDENTIAL DATA FOR RESEARCH  
23 JUNE 2020

SURF

# Barrier for data sharing

⚠ Gain is usually with the data consumer, burden is with the data provider



# Willingness to share data

***ROI***

+

***Trust***



Return on Investment (ROI) is determined by the balance between effort it takes to share data, and the gain received by sharing data

Trust is determined by the balance between the risks (due to privacy or competition), and the control (due to verification and security) of sharing and usage of data

**Gains**

**Effort**



Return on investment

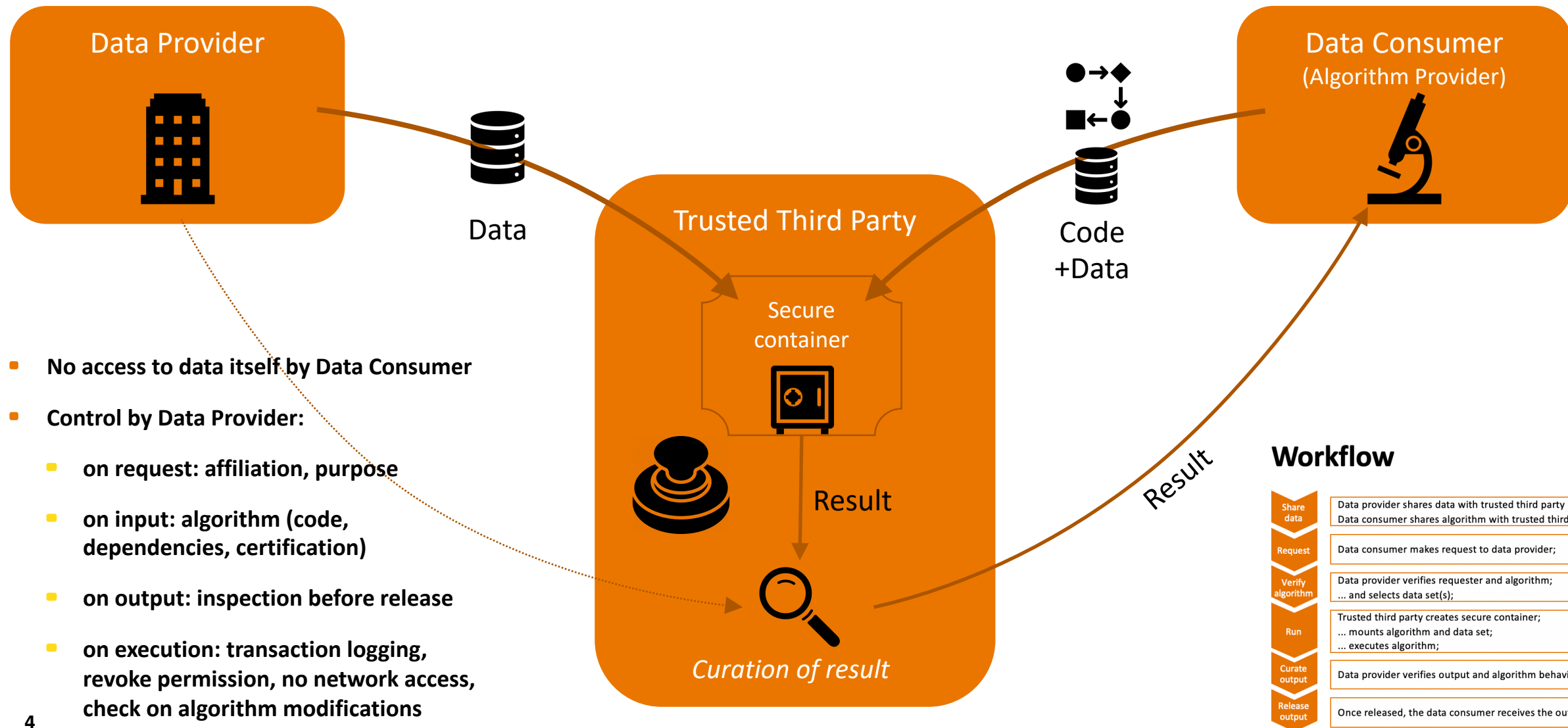
**Control**

**Risk**



Trust

# Working prototype on trusted data sharing



Shared with others - SURF - Re X




https://researchdrive.surfsara.nl/index.php/apps/files/?dir=/&view=sharing

Menu Files

SURF RESEARCH DRIVE

freek@macfreak.nl

- All files
- Favorites
- Shared with you
- Shared with others
- Shared by link
- External storage
- Deleted files
- Settings

| Name   | Share time                             |
|--|--|
|  ugly_cats_and_dogs       | Shared with Data Exchange 2 days ago   |
|  cat-looks-like-dog-1.jpg | Shared with Data Exchange 13 hours ago |
|  random_numbers.txt       | Shared with Data Exchange 2 days ago   |
| 1 folder and 2 files   |  |



Files - SURF - Research Drive My Files

https://researchdrive.surfsara.nl/index.php/apps/files/?dir=/&fileid=81846599

SURF RESEARCH DRIVE

Menu Files Freek Dijkstra

All files Favourites Shared with you Shared with others Shared by link External storage Deleted files Settings

calculate\_sum.py calculate\_average.py demo1\_data demo1\_code Data Exchange ComputeWizards catordog catdog

Name Size Modified

calculate\_sum.py < 1 KB a month ago

calculate\_average.py < 1 KB a month ago

demo1\_data Mike Kotsur < 1 KB 14 days ago

demo1\_code Mike Kotsur < 1 KB a month ago

Data Exchange Data Exchange 0 KB 2 months ago

ComputeWizards Shared 0 KB 3 months ago

catordog Shared 121.8 MB 13 hours ago

catdog Mike Kotsur 121.8 MB a day ago

calculate\_sum.py < 1 KB, a month ago

Activities Comments Sharing

User and Groups Public Links

Data Exchange

6 Algorithm is shared with the Data exchange by researcher



My Files Files - SURF - Research Drive +

< > ↺ | <https://dataexchange.surfsara.nl/tasks/request> 🔒

**DataExchange (Demo)** Home My permissions **Create Request** Log out (freek.dijkstra@surfsara.nl) — Algorithm owner (Toggle)

## Create request

### Request Permission for a dataset

Type of permission

The selected algorithm will be ran on the selected dataset of the data owner exactly once.

Select algorithm

Data owner email

Dataset description

### Run an algorithm with continuous permission

Select algorithm

Select dataset

Shared with others - SURF - Re X My Files

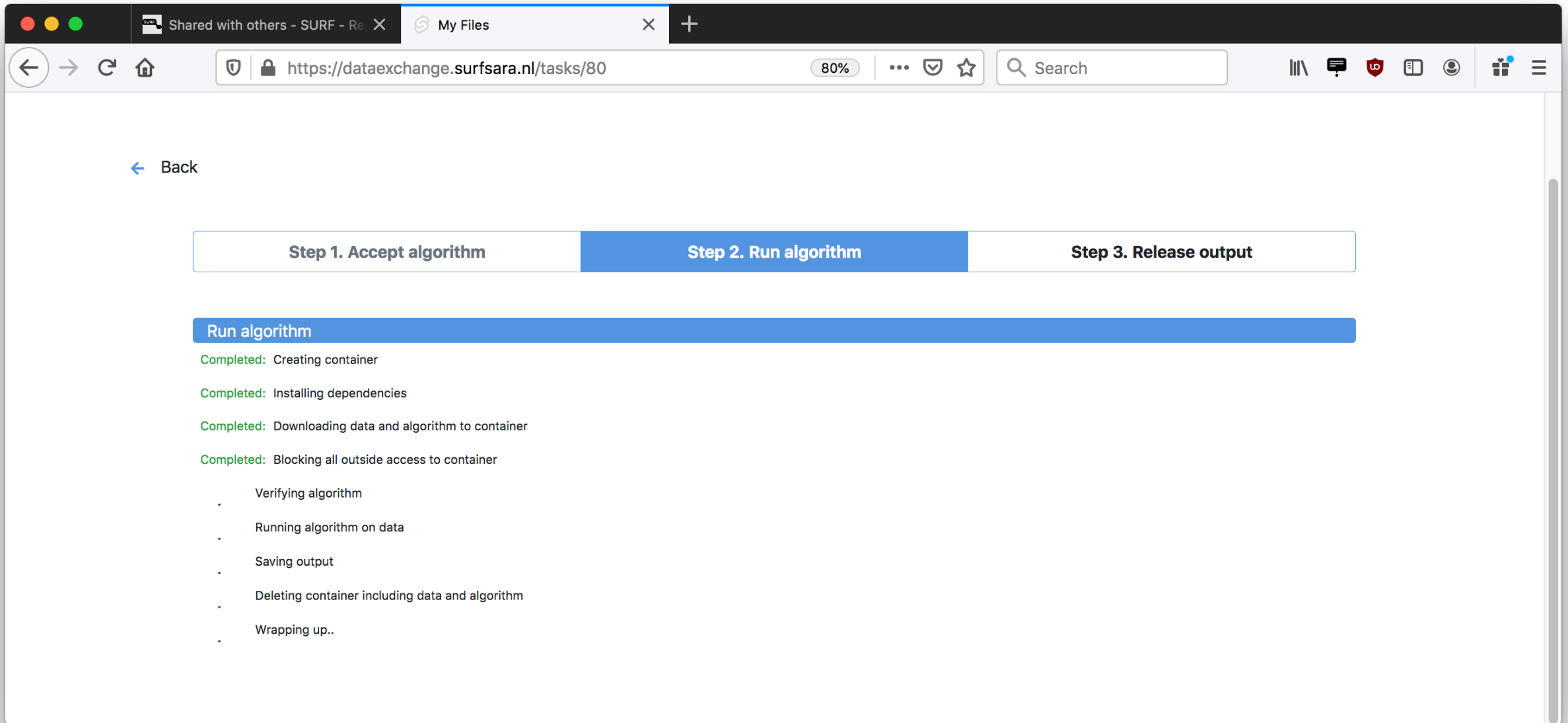
https://dataexchange.surfsara.nl/tasks/80 80% Search

**Step 1. Accept algorithm** Step 2. Run algorithm Step 3. Release output

|  |  |  |
|--|--|--|
| <b>Algorithm Owner</b><br>freek.dijkstra@surfsara.nl   | <b>Algorithm Name</b><br>calculate_sum.py                        | <b>Algorithm Code</b><br><pre>calculate_sum.py #!/usr/bin/env python import sys  def main():     try:         filename = sys.argv[1]     except IndexError as err:         sys.stderr.write("Please spec     return -1      sum = 0     with open(filename, 'r') as numbe         for number in numbersfile:             try:                 sum += int(number)             except ValueError as err:</pre> |
| <b>Permission Type</b><br>one time permission  | <b>Algorithm Dependencies</b><br>sys                             |  |
| <b>Permission Information</b><br>The selected algorithm will be ran on your selected dataset once. | <b>Algorithm Length</b><br>Lines: 22, Words: 44, Characters: 522 |  |
|  | <b>Choose dataset</b><br>random_numbers.txt                      |  |

Run algorithm on data to see output and go to step 2 Reject request





Shared with others - SURF - Re X My Files

https://dataexchange.surfsara.nl/tasks/80 80% Search

Step 1. Accept algorithm Step 2. Run algorithm Step 3. Release output

| Algorithm Owner   | Algorithm Name                        | Output |
|---|---------------------------------------|--------|
| freek.dijkstra@surfsara.nl  | calculate_sum.py                      | 30717  |
| Permission Type   | Algorithm Dependencies                |        |
| one time permission   | sys                                   |        |
| Permission Information  | Algorithm Length                      |        |
| The selected algorithm will be ran on your selected dataset once. | Lines: 22, Words: 44, Characters: 522 |        |
|   | Used Dataset                          |        |
|   | random_numbers.txt                    |        |

Reject output Release output

My Files Files - SURF - Research Drive +

< > ↺ 🗑️ | 🔒 https://dataexchange.surfsara.nl/tasks/80 🔍 📷 ⌘ ❤️ | 🛡️ 🚀 🧑 🌱 ⬆️ ⬇️ ⚙️

Execution finished

| Data Owner   | Algorithm Name                        | Output |
|--|---------------------------------------|--------|
| freek.dijkstra@surfsara.nl   | calculate_sum.py                      | 30717  |
| <b>Permission Type</b>   | <b>Algorithm Dependencies</b>         |        |
| one time permission  | sys                                   |        |
| <b>Permission Information</b>  | <b>Algorithm Length</b>               |        |
| The selected algorithm will be ran on the selected dataset of the data owner exactly once. | Lines: 22, Words: 44, Characters: 522 |        |
|  | <b>Choose dataset</b>                 |        |
|  | random_numbers.txt                    |        |

Shared with others - SURF - Re X Manage Data X +

← → ↺ 🏠 [https://dataexchange.surfsara.nl/manage\\_data](https://dataexchange.surfsara.nl/manage_data) 80% ... 🛡️ ☆ 🔍 Search

**DataExchange (Demo)** Home **Manage Data** Review Requests Log out (freek@macfreek.nl) — Data owner (Toggle)

### Manage shared Files and Folders

📄 random\_numbers.txt **Withdraw Data**

#### Permissions

| With                       | Algorithm        | Type                |                          |
|----------------------------|------------------|---------------------|--------------------------|
| freek.dijkstra@surfsara.nl | calculate_sum.py | one time permission | <b>Reject Permission</b> |

#### Runs

| Algorithm Owner            | Passed   | When       | Action                  |
|----------------------------|----------|------------|-------------------------|
| freek.dijkstra@surfsara.nl | Passed   | 20-11-2019 | <a href="#">See log</a> |
| freek.dijkstra@surfsara.nl | Rejected | 18-11-2019 | <a href="#">See log</a> |

📄 ugly\_cats\_and\_dogs **Withdraw Data**

# Next Steps

- Better understanding of the needs and requirements
  - This webinar!
- Work with potential pilot partners
  - Integrate with ODISSEI Data Node
  - Talks with other interested organisations (perhaps you?)
- Make the demo accessible to everyone.
  - Currently requires a SURF ResearchDrive account
  - We'll make it work with e.g. Google Drive (August 2020)

# Different Methods to Ease Data Sharing

## Agreements

- Stipulation of what can/cannot be done
- Signing of contract or NDA
- Dispute resolution process

## Registration

- Authentication
- Verification of credential
- Reputation score
- Policy framework
- Audit trails

## Pseudonymization

- Filtering (on records)
- Pruning (on properties)
- Aggregation (combine records)
- Make coarse grained buckets
- Slight alteration of data
- One-way hashing
- One-time identifiers
- Synthetic data (mix records / AI)

## Data Vault

- Data source retains control
- Delegate permissions
- No central data lake
- Data marketplace

## Secure Containers

- Bring algorithm to data
- At Trusted third party or at data provider
- Share output instead of data

## Secure Computing

- Secure multi-party computation
- Homomorphic encryption
- Garbled Circuits
- Zero-knowledge proof



# COLLABORATION WITHOUT SHARING DATA



Freek Dijkstra



Freek.Dijkstra@surfsara.nl



www.surf.nl



This presentation is available under the  
creative commons attribution 4.0 license

## Driving innovation together

SURF